



ORACLE®

The Cloud, Some Big Data, and Research: an Example

Harry J Foxwell, PhD
Principle Consultant for Cloud Computing
Oracle Public Sector
harry.foxwell@oracle.com

National Cancer Institute & Oracle Cloud / Big Data Project

- **Winner: Best Federal Big Data Project at Hadoop World, NYC, 2012:**
 - <http://smartdatacollective.com/bobgourley/86241/winner-2012-government-big-data-solutions-award-national-cancer-institute>
- **Studied relationship between genes & cancers:**
 - built infrastructure to cross-reference relationships between 17000 genes and five major cancer subtypes across *20 million biomedical publication abstracts*
 - cross referenced TCGA gene expression data from *simulated 60 million patients* and miRNA expression for *a simulated 900 million patients*

Cloud-based Data Sources

National Cancer Institute

at the National Institutes of Health



genome.gov

National Human Genome Research Institute

National Institutes of Health



U.S. National Library of Medicine

National Institutes of Health

PubMed.gov

US National Library of Medicine
National Institutes of Health

The Cancer Genome Atlas



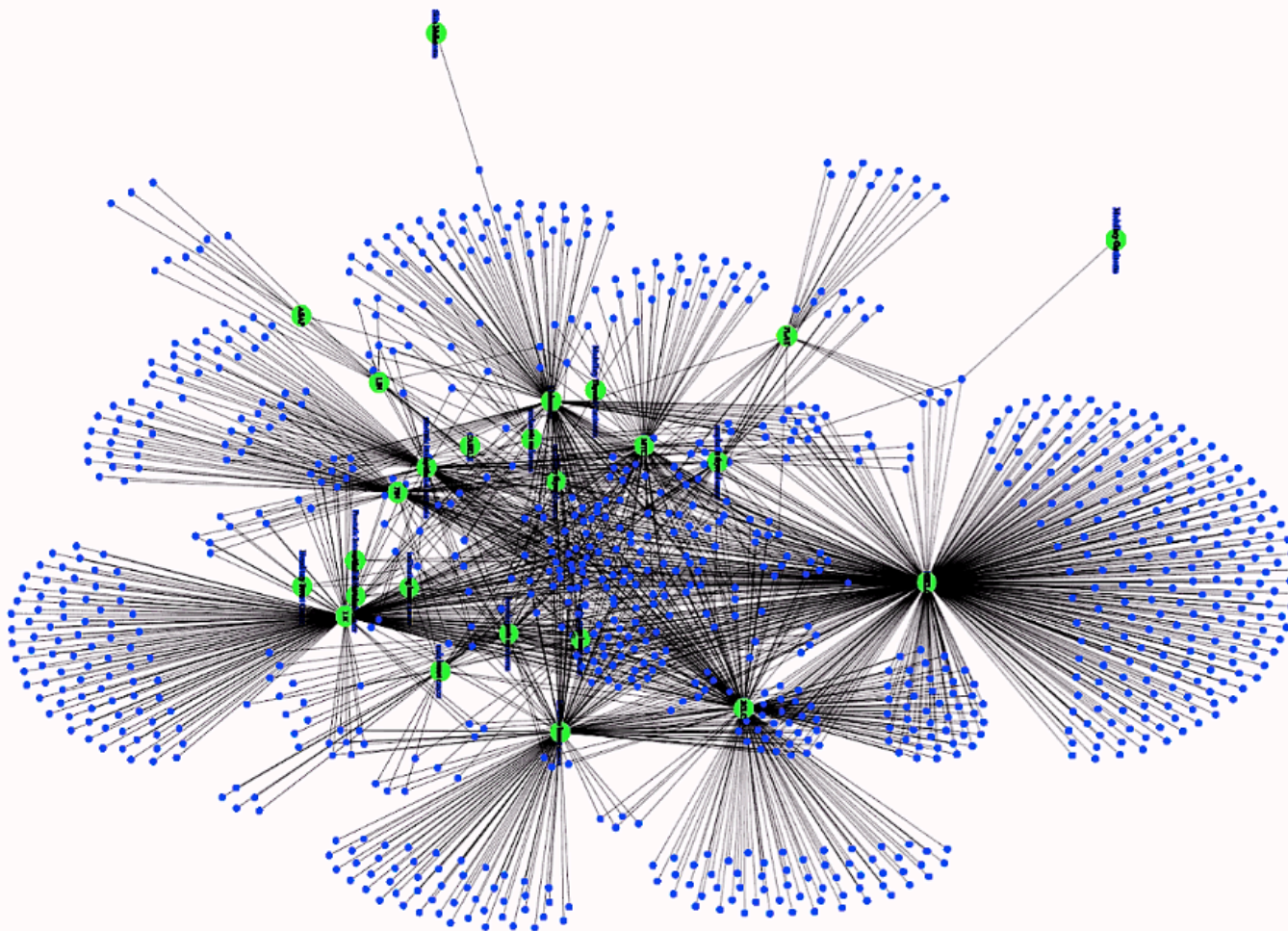
*Understanding genomics
to improve cancer care*

What does this mean?

- **Given a human's gene sequence:**
 - Identify those genes that have been associated with a cancer
 - Correlate that gene to known cancer types by identifying NIH research publications which reference this gene and cancer type
 - Present the information graphically



Results (NCI) – Cancer Co-Occurrence



Insights from the Analysis

- Used open source Hadoop/HDFS, Apache Hive
- Notice the number of genes associated with some cancers - some have a lot; others very few
- 1: Size of bubble is # of publications (proxy for \$\$)
- 2: Top 20 genes: some genes shared across multiple cancer types, others unique to particular cancer
- Data ingest took longest...then **queries took less than 1 minute to complete!**

oracle.com/cloud



www.facebook.com/OracleCloudComputing



[@OracleCloudZone](https://twitter.com/OracleCloudZone) Hashtag: [#oraclecloud](https://twitter.com/hashtag/oraclecloud)